# Manual instalação Cloudera Data Science Workbench 1.5 – CentOS 7.5

## Pré- requisitos pra instalação

|  |  |
| --- | --- |
| **Operating System** | **Versions** |
| RHEL / CentOS / Oracle Linux RHCK | 7.2, 7.3, 7.4, 7.5 |
|
|
| Oracle Linux (UEK - default) | 7.3 |
| SUSE Linux Enterprise Server (SLES) | 12 SP2, 12 SP3 |

|  |  |  |
| --- | --- | --- |
| **Resource Type** | **Master** | **Workers** |
| CPU | 16+ CPU (vCPU) cores | 16+ CPU (vCPU) cores |
| RAM | 32+ GB | 32+ GB |
| **Disk Space** | | |
| Root Volume | 100+ GB | 100+ GB |
| Application Block Device | 1 TB | - |
|
| Docker Block Device | 1 TB | 1 TB |

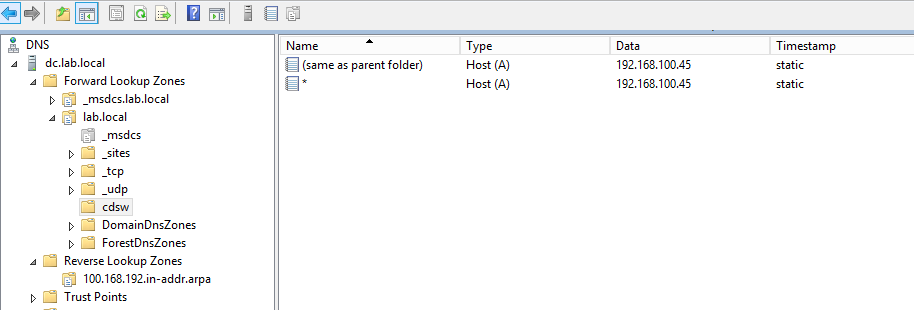
<https://www.cloudera.com/documentation/data-science-workbench/latest/topics/cdsw_requirements_supported_versions.html>

## Configurar DNS

Configurar no servidor de DNS duas entradas conforme informações e imagem abaixo:

cdsw.node4.com. IN A 192.168.100.45

\*.cdsw.node4.com. IN A 192.168.100.45



*To set up subdomains for Cloudera Data Science Workbench, configure your DNS server with an A record for a wildcard DNS name such as \*.cdsw.<your\_domain>.com for the master host, and a second A record for the root entry of cdsw.<your\_domain>.com.*

Considerando que o JAVA HOME já estiver setado e configurado corretamente, poderá seguir o passo a passo a abaixo

## **Download arquivo jar**

Download do arquivo jar e colocar no caminho /opt/cloudera/csd do node máster do Cloudera manager server.

<https://archive.cloudera.com/cdsw1/1.5.0/csd/CLOUDERA_DATA_SCIENCE_WORKBENCH-CDH5-1.5.0.jar>

## **Mudar o dono do arquivo**

chown cloudera-scm:cloudera-scm CLOUDERA\_DATA\_SCIENCE\_WORKBENCH-CDH<X>-1.5.<Y>.jar

## **Mudar as permissões do arquivo**

chmod 644 CLOUDERA\_DATA\_SCIENCE\_WORKBENCH-CDH<X>-1.5.<Y>.jar

## **Restartar o serviço do cloudera server**

service cloudera-scm-server restart

## **Restartar o cloudera management service**

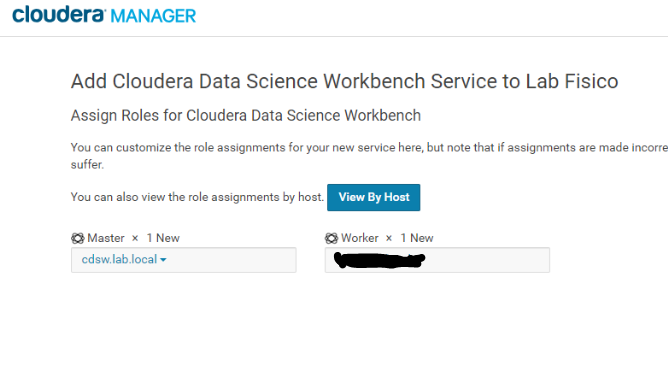
Entrar no Cloudera Manager Admin console -> Cluster -> Cloudera Management Service -> Actions -> Restart

## **Iniciar o Parcel no CDH**

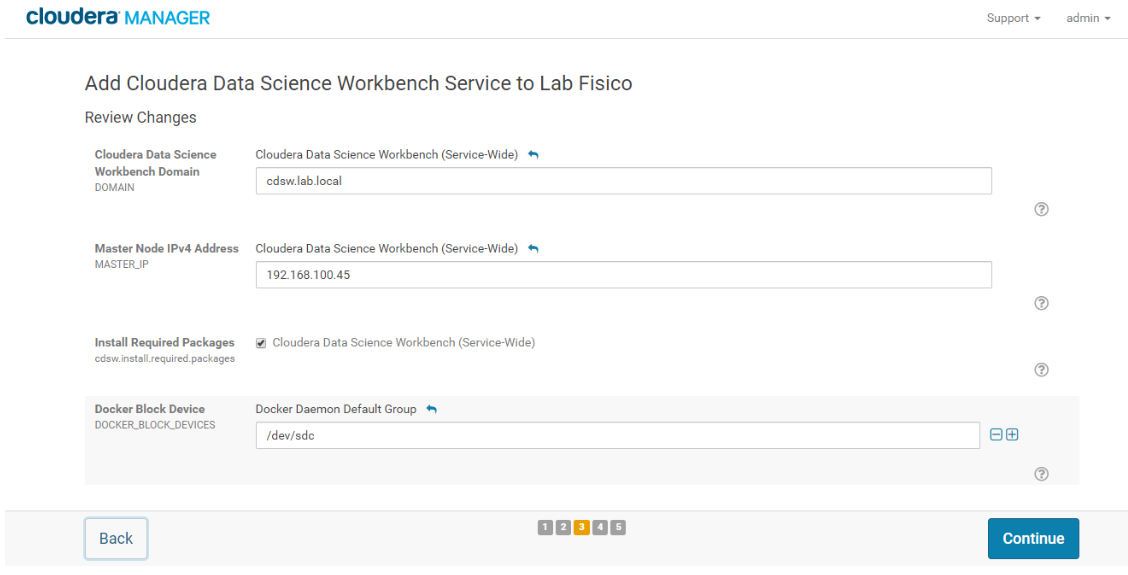
Entrar no cloudera manager ir para o menu Hosts -> Parcels -> CDSW -> Download -> Distribute -> Active

## **Adicionar o serviço do CDSW**

Logar no cloudera manager e ir em adicionar serviço. Adicionar o serviço CDSW, clicar em continuar.



Se estiver instalando em apenas um node o CDSW, não é necessário preencher a seção do worker node. Clicar em continuar



Continuar e finalizar.

## **Verificar o arquivo cdsw.conf**

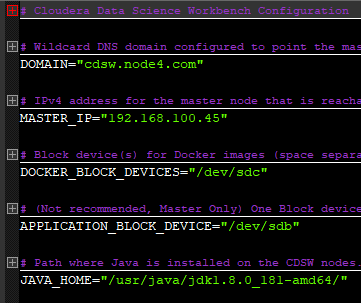
Uma boa prática é verificar as informações do arquivo cdsw.conf. para o correto funcionamento do CDSW.

Provável caminho para localização do arquivo: /etc/cdsw/config/cdsw.conf

Exemplo de arquivo cdsw.conf já configurado



<https://www.cloudera.com/documentation/data-science-workbench/latest/topics/cdsw_install_package.html>

Campos obrigatórios para alteração

[https://www.cloudera.com/documentation/data-science-workbench/latest/topics/cdsw\_install\_parcel.html#add\_cdsw\_service](https://www.cloudera.com/documentation/data-science-workbench/latest/topics/cdsw_install_parcel.html" \l "add_cdsw_service)

## **Status do serviço CDSW**

*cdsw status*

## **Visualizar a situação do serviço**

A aplicação leva alguns minutos para subir, mas é possível visualizar a situação do serviço através do comando:

*watch cdsw status*

## **Observações**

A instalação normalmente leva em torno de 30 minutos, mas isso poderá se postergar dependendo se quantos máquinas serão instaladas o CDSW e também quais os pacotes necessários para cada máquina, exemplos de pacotes – R, Python, Scala

Depois de finalizar a instalação com sucesso é necessário configurar a conta de administrador, vá até ao acesso via web http://cdsw.<company>.com

## **Erros conhecidos no cdsw status**

Error in checking sysctl params:

ERROR:: net.bridge.bridge-nf-call-ip6tables is not set to 1: 1

## **Solução**

Adicionar as seguintes linhas no /etc/sysctl.conf

add

net.bridge.bridge-nf-call-iptables=1

net.bridge.bridge-nf-call-ip6tables=1